D1-82-0252

BOEING SCIENTIFIC RESEARCH LABORATORIES

A Statistical Analysis of the Evidence of the Beneficial Effect of the Use of an Estrogen in Prostatectomy

Sam C. Saunders

Mathematics Research

June 1963

D1-82-0252

A STATISTICAL ANALYSIS OF THE EVIDENCE OF THE BENEFICIAL EFFECT OF THE USE OF AN ESTROGEN IN PROSTATECTOMY

by

Sam C. Saunders

Mathematical Note No. 297

Mathematics Research Laboratory

BOEING SCIENTIFIC RESEARCH LABORATORIES

June 1963

PREFACE

This note was the result of my collaboration with C. D. Goodhope, M. D. to remove the difficulties in using a classical t-test of the hypothesis of the equality of two means when applied to the bloodloss during surgery of patients treated or in control. The t-test assumes the underlying distributions are normal or nearly so. A glance at the data indicated that this was an untenable assumption since several observations were too many standard deviations away from the mean. Consequently, a less restrictive non-parametric method was called for.

In order to explain the rationale of this analysis, and to preserve some of the things that we have mutually learned from our discussions, this note was written. It shows again that statistical techniques which apply to mechanical machines will apply to human ones as well.

SUMMARY

In this paper we examine some data which purports to show that patients undergoing prostate surgery can be protected from excessive bloodloss through the prior administration of an estrogen. We point out some assumptions that must be made in order to assure that a test of measured bloodloss is equivalent with a test of the actual bloodless these quantities differing by a random error.

We conclude that the use of an estrogen is effective in reducing bloodloss at the 5 per cent significance level but we have insufficient evidence to conclude that it reduces the length of time necessary for the operation. The treatment does seem to reduce the weight of the tissue removed at the 20 per cent significance level.

We also point out that a comparison of average bloodless per minute of operating time probably does not give any additional information than a comparison of bloodloss.

1: Introduction

The population of males who will undergo transurethral prostatectomy during the next few years is the group about which our inferences are to be made. To do so we shall use the clincal data that was obtained by C. D. Goodhope [1]. Thus, strictly speaking, the subpopulation from which our sample has been taken is that of the males in a particular geographical area who have presented themselves for clinical treatment to one urologist during the last few years. However we have no reason to believe that this subpopulation is not representative of the population with which we are concerned.

Some of the random variables (recall a random variable is any quantitative measurement made on a member of a population) which we shall consider are the bloodloss during surgery, the weight of the tissue removed, and the length of time of the operation.

We wish to compare the effect of a particular treatment on these random variables. Since the difference of these quantities between members of the population is quite sizable even though both were accorded the same care throughout, a meaningful comparison of the effect of the treatment, if there is any, can only be made through statistical means. Specifically we shall test for statistical significance the effect of 40 mg of polyestradial phosphate (P.E.P.) in 2 c.c. solution administered intramuscularly one week prior to surgery against that of placebo (2 c.c. saline solution similarly administered) by means of statistical tests at a specified level of confidence.

It is anticipated that the effect of this treatment is, in general, to reduce the actual bloodloss as well as make the operation easier to perform by reducing the weight of the tissue being removed and thus consequently to reduce the length of time necessary for the operation.

In this paper we shall not be concerned with the chemical mechanism whereby the estrogen P.E.P. produces its effect. In this regard one is referred to the publications of its discoverer Diczfalusy [2] and to that of Goodhope [1], as well as the references given there. Rather we shall be concerned with the fulfillment of the requirements of correct statistical analysis and the relationships between the mathematical assumptions and actuality.

2: The Data and Its Collection

From Goodhope [1], we have the following data:

	CONTROLS						ESTRADURIN PROTECTED				
PT.	WT. TISSUE GMS.	TOTAL BLOOD LOSS CC	LOSS PER GM.	OP TIME MIN.	LOSS PER MIN.	PT.	WT. TISSUE GMS.	TOTAL BLOOD LOSS CC	LOSS PER GM.	OP TIME MIN.	LOSS PER MIN.
EP	25	308	12.3	40	7.7	WA	9	20	2.2	30	.66
FW	11	780	70.9	30	26.0	AP	6	197	32.8	40	4.8
JN	8	108	13.6	24	4.5	JF	5	20	4.	17	1.2
CG	24	180	7.5	40	4.5	HT	23	38	1.6	50	.76
CR	25	329	13.2	62	5.3	EH	10	30	3.	45	.66
*GP	40	785	19.6	100	7.85	*GP	22	65	2.9	75	.86
WW	7	300	42.8	47	6.4	TS	17	310	18.2	50	6.2
JE	9	250	27.7	40	6.25	VL	20	110	5.5	25	4.4
AVE.	18.62	392.5	25.95	48	8.56		14.0	98.75	8.77	41.5	2.45

It was obtained from sixteen cases (fifteen patients - one was operated on twice once under control and once under treatment) suffering from prostatic obstruction and operated by transurethral prostatectomy. Now the measured

bloodless, the length of time for the operation, and the weight of the tissue excised for each patient under the surgery described are the observed values of the random variables in which we are interested. We want to determine the effect of treatment on these random variables. We shall be concerned with several statistics also, principally the ones listed.

To observe a value of the statistic for the control or treated group we obtain a patient with prostatic obstruction and then decide by a chance mechanism with probability one half whether he will be in the treated group or not, the next patient being in the alternate group. The operation is performed and we determine the measured bloodloss, the duration of the operation, and the weight of the tissue excised for that patient.

We assume:

1°: The statistics for the control and treated groups are independent random variables each measurement of which is continuous.

Independence means that a knowledge of the statistic for a control patient will not provide any information as to the outcome of the statistic for a treated patient and vice versa. The guaranteeing of independence is the raison d'etre of the selection procedure and the performance of hemotological studies to exclude abnormal bleeding or clotting tendencies in every case.

The assumption of continuity means that the methods of measurement is of sufficiently refined scale that it is always possible to determine between two patients which measurement has the greater magnitude, that

is we cannot have a tie between patients in this respect. This can be effected by the choice of meterology. The only case where the meterology should be mentioned is bloodloss. Measurement was accomplished by the determination of total hemoglobin content in all washings by colorimetric methods. Test indicate that the accuracy of this method was within $\pm$ 15 per cent for small losses and $\pm$ 5 per cent for large losses. However, we cannot assume that this implies that the measured bloodloss is within such percentages of the true bloodloss because in certain cases blood may coagulate in the washings and thus not be measured. Also in other cases clotting may take place in the bladder and therefore not be measured. These occurrences are presumed to be more probable with the greater loss of blood.

3: The Actual Bloodloss and the Measured Bloodloss

Let M_1 be the controlled variable measured bloodloss and M_2 the treated variable measured bloodloss. The probability law that governs M_1 we shall denote by F_1 with the interpretation

$$\text{Prob}\{M_1 \leq a\} = F_1(a) \quad \text{for each real } a > 0,$$

and similarly we shall assume M_2 has a distribution law F_2 and

$$\text{Prob}\{M_2 \leq a\} = F_2(a) \quad \text{for each real } a > 0.$$

We want to test the hypothesis that M_1 is equal to M_2 in the stochastic sense, which is to say these distributions are equal,

$$H_0 : F_1 = F_2$$

against the alternative that M_2 is stochastically smaller than M_1, i.e.

$$H_1 : F_1 < F_2.$$

This may seem to be confusing at first thought, but it is precisely what is needed. Under the null hypothesis H_0 there is no difference between the treated group and the control group, which is to say that

$$\text{Prob}\{M_1 \leq a\} = \text{Prob}\{M_2 \leq a\} \quad \text{for every} \quad a > 0.$$

In words: The probability that a treated patient will lose an amount of blood not exceeding a units is exactly equal to the probability an untreated patient will lose the same amount, moreover this equality holds for all amounts of blood a.

Under the alternative hypothesis that the treatment is effective, we have

$$\text{Prob}\{M_2 > a\} < \text{Prob}\{M_1 > a\} \quad \text{for every} \quad a > 0.$$

In words: The probability that a treated patient will lose more than a units of blood is less than the probability that an untreated patient will lose more than the same amount a and we have the inequality holding no matter what this amount is.

A central question at this juncture is whether a test of an hypothesis on the distribution of the measured bloodloss M is what one is really interested in. Of course, the answer is "no." We are primarily interested in the distribution of the true bloodloss and it is this random variable and its distribution about which the hypothesis should be made. Unfortunately, the true bloodloss of a patient we can never know. The measured bloodloss is always different by a meterology error (which we assume is

random). Letting the true bloodloss, in either case, treatment or control, be L and the measured bloodloss be M_L, we have

$$M_L = L + \epsilon_L$$

where ϵ_L is the random error due to the accuracy of the colorimeter when measuring a true bloodloss of amount L as well as the amount of blood lost which is not measurable for one reason or another. As previously noted the distribution of error depends upon the true bloodloss.

We now make the assumption

2° : Given that the true bloodloss will be a certain amount, the meterology error will have the same distribution for a patient with treatment as one without. That is

$$P[M_\ell \leq x | L_i = \ell] = J(x|\ell) \qquad \text{for } i = 1,2$$

and J does not depend upon i.

3° : As the given total bloodloss increases the random measured bloodloss increases in a stochastic fashion. That is for each fixed x, $J(x|\ell)$ is a decreasing function of ℓ. We further assume its derivative with respect to ℓ exists and is continuous and not zero at $x = \ell$.

Under control we let the actual random bloodloss be L_1 with distribution G_1, under treatment let actual random bloodloss be L_2 with distribution G_2. In all cases we have $G_1 \leq G_2$ and we describe two states of nature

$$H_0' : G_1 = G_2 \qquad\qquad H_1' : G_1 < G_2 .$$

We now prove a result which says that from 2° and 3° it follows that a test of H_0 against H_1 is equivalent with a test of H_0' against H_1'.

We have the

Theorem 1: Under the assumptions $G_1 \leq G_2$ and $F_1 \leq F_2$ with 2° and 3° holding we have

$$F_1 = F_2 \text{ iff } G_1 = G_2$$

$$F_1 < F_2 \text{ iff } G_1 < G_2 \ .$$

Proof: By the theorem of total probability we have for $i = 1,2$

$$F_i(x) = P[M_{L_i} \leq x] = \int_0^\infty J(x|\ell)G_i'(\ell)d\ell \ .$$

Integration by parts yield, upon setting $v(x|\ell) = -\frac{d}{d\ell} J(x|\ell)$

$$(*) \qquad F_i(x) = J(x|\infty) + \int_0^\infty G_i(\ell)v(x|\ell)d\ell$$

and by 3° we have that $J(x|\ell)$ is a decreasing function of ℓ for fixed x and thus $\lim_{\ell\to\infty} J(x|\ell)$ exists and equals say $J(x|\infty)$ moreover $v(x|\ell)$ is always nonnegative.

Thus by (*) it is clear that if $G_1 = G_2$ then $F_1 = F_2$ but also if $F_1 = F_2$ then for every x we must have

$$\int_0^\infty [G_2(\ell) - G_1(\ell)]v(x|\ell)d\ell = 0.$$

Since $G_2 \geq G_1$ and $v(x|\ell)$ is nonnegative we must have

$$G_2 = G_1 \qquad \{\ell : v(x|\ell) > 0 \quad \text{for some } x\}$$

but by the assumption in 3° this is the entire space $\{\ell > 0\}$.

Remark: Assumption 3° would for example follow if ϵ_ℓ had a distribution say J independent of ℓ which would mean that the error would be stochastically of the same magnitude no matter what the true bloodloss would be. In this case we would have $J(x - \ell) = J(x|\ell)$.

Based upon the reasonableness of assumptions 2° and 3° the theorem above proves that the performance of a statistical test about the distribution of the measured bloodloss is the same as the performance of a test on the distribution of the true bloodloss of patients with and without treatment. This fact is of primary importance. Merely because it is overlooked so often should not blind us from seeing its neglect can often vitiate the conclusions reached.

4: Analysis of the Data: Operating time, Weight of excised tissue, and Bloodloss

In this section we shall make use of the Wilcoxon-Mann-Whitney statistic which is

$$U = mn + m(m + 1)/2 - S$$

where S is the sum of the ranks of the controlled variates and m and n are the number of observations of control and treatment, respectively. Probability tables of the distribution of U are given in [3].

Operating Time in Minutes

Control		24		30		40	41	42			47			62		100
Treated	17		25		31				43	45		50	51		75	
Rank		2		4		6	7	8			11			14		16

$$U = 100 - 68 = 32$$

which is not significant. In this case, under the null hypothesis,

$P[U \leq 32] = .520$. Therefore we cannot reject the null hypothesis that the operating time of treated patients is less than the operating time of untreated patients.

Weight of Excised Tissue in Grams

Control			7	8	9			11					24	25	25+	40
Treated	5	6				9+	10		17	20	22	23				
Rank			3	4	5			8					13	14	15	16

$U = 22$.

This value of U is significant at the 20 per cent level. In this case $P[U \leq 22] = .191$ under the null hypothesis. Thus it does appear reasonable that later observations may prove that the weight of excised tissue is reduced significantly.

Measured Bloodloss in Cubic Centimeters

Control						108		180		250	300	308		329	780	785
Treated	19	20	30	38	65		110		197				310			
Rank						6		8		10	11	12		14	15	16

$U = 8$.

This value of U is significant at the 1 per cent level. Under the null hypothesis $P[U \leq 8] = .005$. Hence we reject the hypothesis that there is no difference in bloodloss between the control and treated group and we do so at the 1 per cent level of significance. We conclude that the treatment is effective.

One might think that a comparison of bloodloss per minute between the two populations would be informative since one notes that in one case the average is nearly three times that of the other. However, it does not give

us additional information since a test of bloodloss per minute is the same as a test of measured bloodloss under some reasonable assumptions. These assumptions are

4°: Given that the bloodloss is the same for a treated and an untreated patient the times for the operations would be stochastically equal.

5°: Given the true bloodloss the error in measured bloodloss is independent of the length of time for the operation.

6°: As the true bloodloss increases the average measured bloodloss per minute increases in a stochastic manner.

In view of the statistical nonsignificance of the time length of the operation between the treated and the control group even when disregarding the variation of total bloodloss, which is some measure of the difficulty of the operation and hence of its duration, we feel that 4° is reasonable.

Assumption 5° seems undeniable. Assumption 6° is not so intuitive, however, a glance at the data seems to substantiate this assumption. We also present the following reasoning: In the case of active bleeding the surgeon spends more operating time than otherwise in effecting coagulation to reduce the bleeding rate to an acceptable level. However, it is clear that in the case of hypothetical active bleeding at a constant rate the average true bloodloss will be higher than if the bleeding rate were constant at an acceptable rate. The extra time spent in coagulation cannot reduce the average bloodloss.

However, regardless of the profile of the patients bleeding rate during the operation it is still true that extra time spent in efforts to inhibit

bleeding cannot reduce the average bloodloss per minute if at the time coagulation efforts are begun the average of the instantaneous acceptable and unacceptable rates exceed the average per minute loss at the acceptable rate.

To see this consider a safe bleeding rate r_a for an operation so that during a normal operation of length T the average bloodloss per minute would be

$$\frac{1}{T}\int_0^T r_a(t)dt = \bar{L} \text{ say.}$$

An operation is begun which has an unacceptable bleeding rate, say r_u, and at time τ the surgeon observes that the average of the acceptable and unacceptable rates exceeds $\bar{L}$, i.e.

$$\frac{1}{2}\left(r_u(\tau) + r_a(\tau)\right) \geq \bar{L} .$$

We then begin coagulation efforts reducing the bloodloss linearly to an acceptable level r_a in a period of length c minutes and then proceeds with the remainder of the operation taking $(T - \tau)$ minutes we then would have

$$\bar{L} \leq \frac{1}{T+c}\left[\int_0^\tau r_u + \frac{r_u(\tau) + r_a(\tau)}{2}\, c + \int_\tau^T r_a\right] .$$

This is equivalent to

$$(T + c)\int_0^T r_a(t)dt \leq T\int_0^\tau r_u(t)dt + \frac{cT}{2}\left[r_u(\tau) + r_a(\tau)\right] + T\int_\tau^T r_a(t)dt$$

$$c\int_0^T r_a(t)dt \leq T\int_0^\tau [r_u(t) - r_a(t)]dt + \frac{cT}{2}\left[r_u(\tau) + r_a(\tau)\right]$$

$$\bar{L} \leq \frac{1}{c}\int_0^\tau [r_u(t) - r_a(t)]dt + \frac{[r_u(\tau) + r_a(\tau)]}{2}$$

which proves the contention.

We now have the

Theorem 2: Let R_i be the distribution of the ratio (M_i/T_i), the average measured bloodloss per minute for $i = 1,2$ then from 4° and 5° we have

$$R_1 = R_2 \quad \text{if} \quad G_1 = G_2 .$$

But further we have

$$R_1 < R_2 \quad \text{if} \quad G_1 < G_2$$

and 6° holds.

Proof: By definition

$$R_i(x) = P[(M_i/T_i) \leq x] = \int_0^\infty P\left[\frac{M_\ell}{T_i(\ell)} \leq x \,|\, L_i = \ell\right] dG_i(\ell)$$

where $T_i(\ell)$ is the (random) time of the operation given the total bloodloss is ℓ which by 4° has a distribution $W(\cdot|\ell)$ which does not depend upon i.

Now we examine the integrand which is

$$P[M_\ell \leq x \cdot T_i(\ell) \,|\, L_i = \ell] = \int_0^\infty P[M_\ell \leq x \cdot t \,|\, L_i = \ell]\; W(dt|\ell)$$

and by 5° we have the equality

$$= \int_0^\infty J(xt|\ell) W(dt|\ell)$$

which does not depend upon $i = 1,2$. Moreover this quantity is by 6° a decreasing function of ℓ. The remainder of the proof is analogous to that given for the previous theorem and need not be given here.

However, the hypothesis that measured bloodloss per gram of excised tissue is stochastically smaller in the treated case is in fact stronger than that the measured bloodloss is stochastically smaller itself in the treated case.

Let V_1, V_2 be the measured bloodloss per gram of severed tissue and W_1, W_2 the weight of the tissue in the control and treated cases respectively. Then if W_2, V_2 are stochastically not greater than W_1 and V_1 respectively, then $M_2 = V_2W_2$ is stochastically not greater than $M_1 = V_1W_1$.

To see this consider the string of inequalities

$$P[V_1W_1 \leq x] = \int_0^\infty F_{V_1}(x/t)\,d\,F_{W_1}(t) \geq \int_0^\infty F_{V_2}(x/t)\,d\,F_{W_1}(t)$$

$$= \int_0^\infty F_{W_1}(x/t)\,d\,F_{W_2}(t) \geq \int_0^\infty F_{W_2}(x/t)\,d\,F_{V_2}(t) = P[V_2W_2 \leq x].$$

Thus we have that $P[M_1 \leq x] \geq P[M_2 \leq x]$ and this is the result desired, recalling the definition of stochastic inequality.

Before we conclude we remark that we have continually laid stress upon hypotheses which were in terms of the distribution functions and not in terms of say the mean loss of blood. The reason for this is two-fold. Firstly, it is conceivable that the treatment could on the average have a mean loss which was less but have a greater percentage of patients which require transfusions, and secondly, it is clear from the data that it is not from a normal population so that conclusions via the t-test would not be strictly usable.

BIBLIOGRAPHY

[1] C. D. Goodhope, The Effect of an Estrogen on Bloodloss at Prostatectomy, The Journal of Urology, Vol. 84, pp. 386-389, (1960).

[2] E. Diczfalusy, Polyestradial Phosphate: A Long-Acting Water-Soluble Estrogen, Endocrinology, Vol. 54, pp. 471-477, (1954).

[3] H. B. Mann and H. R. Whitney, On a Test of Whether One of Two Random Variables is Stochastically Larger than the Other, Ann. of Math. Stat., Vol. 18, pp. 50-60, (1947).